Packetized-Voice/Data Integrated Transmission on a Token Passing Ring Local Area Network*

Technical Report 86-22

Tatsuya Suda
Tracy T. Bradley

Department of Information and Computer Science
University of California, Irvine
Irvine, CA 92717
U.S.A.
(714) 856-5474

* This material is based upon work supported by the National Science Foundation under Grant No. DCI-8602052. This research is also in part supported by University of California MICRO program and by the University of California, Irvine, the Academic Senate Committee on Research.

Abstract

This paper investigates the performance of a token passing ring network with packetized-voice/data mixed traffic through extensive simulations. Both data and voice users are modeled in the simulations. Data users produce bursty traffic. Voice traffic is modeled as having alternating talkspurts and silences, with generation of voice packets at a constant rate during talkspurts and no packet generation during silence periods.

The network performance measures obtained include: the distribution of transmission delays for voice packets, the average transmission delay and loss probabilities for voice packets, the number of voice users allowed on a network while satisfying the real-time constraints of speech, and the average transmission delay for data packets.

Token passing ring local area networks are shown to effectively handle both voice and data traffic. The effects of system parameters (e.g., voice packet length, talkspurt/silence lengths, data traffic intensity, and limited versus exhaustive service disciplines) on network performance are discussed.

1. Introduction

In the past few years, local area networks for the interconnection of computers and shared resources within a small area (e.g., a building or a campus) have rapidly gained in popularity. Many architectures have been proposed for local networks, and several have been implemented. These architectures include, among others: a token passing ring [1], a slotted ring [2], and a register insertion ring [3] for ring networks; an Ethernet [4], and a token bus [5] for shared bi-directional bus networks; and a Fasnet [6], and an Expressnet [7] for shared uni-directional bus networks.

The token passing ring network is one of the most popular local networks. Various studies on its performance have established its merit for a variety of data transfer applications [8, 9]. As a rule of thumb, a token passing ring network is less sensitive to changes in network loads, offers reasonably short transmission delays under light loads,[1] and carries more traffic under heavy loads, than other network designs. It also gives less variance in transmission delays than contention type networks, where a large variance in delays is unavoidable due to the undeterministic nature of contention and collision/backoff.

Recently, interest has focused on the use of token passing ring networks for real-time voice applications. These networks provide both computer and telephone communication support for network users [10, 11]. Such integrated service environments have the potential to dramatically increase the productivity of, say, office systems.

The traffic characteristics and performance requirements for voice and data are quite dissimilar [12, 13]. Data traffic can be categorized into two basic types: interactive and bulk data. Interactive data traffic is bursty in nature; the actual proportion of bandwidth utilized is typically very small. It consists of short messages requiring small network delays. Bulk data on the other hand, consists of long messages and requires high throughput, but real-time data delivery is not of primary importance. Strict error control and recovery procedures are required for both types of data.

Voice calls last for a few minutes, and for more than 60% of the conversation time the channel remains idle. This results because only one speaker is active at any particular time; furthermore, there are pauses between sentences, phrases, and even between syllables [14]. Voice traffic can be modeled as having alternating talkspurts and silences, with the generation of voice packets at a constant rate during talkspurts and no packet generation during silences [12, 13, 15]. To preserve the integrity of a conversation, voice packets must be delivered within some time bound (typically 150-200 msec); and, in many data applications, a delay of as much as 200 msec does not present a problem. Conversation is inherently robust, and as a result, speech can be reconstructed at the destination user with acceptable quality, provided that the voice packet loss is less than some specified fraction (typically 1%) [16].

1 It is generally true that contention type networks, such as an Ethernet, give less delay in light load circumstances.

Despite its importance, little has been established about the performance of a token passing ring network when it is subjected to a voice load in addition to a data load. Past research has focused on analyzing its performance for data applications. Hence, this paper evaluates the performance of a token passing ring network through extensive simulations. It examines the dependencies of network performance on both voice and data traffic, and the ways in which system parameters (e.g., voice packet length, data packet length, the number of users) may be altered to allow a network to yield an acceptable level of performance. The performance measures obtained include: the distribution of transmission delays for voice packets, the average transmission delay and loss probabilities for voice packets, the number of voice users allowed on a network while satisfying the real-time constraints of speech, and the average transmission delay for data packets.

As we will see in the following section, all papers on voice/data mixed traffic in the local area network context obtain network performance results through simulations, experimental tests, or approximate analysis; there are no realistic analytic results available for packetized voice transmission. This is due to the correlation in voice packet generation which makes analysis of queueing processes quite difficult. Thus, we take the simulation approach in this paper.

Previous research related to voice and data transmission on local area networks is discussed in Section 2. The network protocols and voice/data traffic models assumed in our simulations are described in Section 3. Section 4 discusses some performance measures (e.g., delay and loss probability) for a packetized voice network. Numerical results and discussions are presented in Section 5.

2. Related Work

The feasibility of packetized voice/data transmission was first discussed in the networking literature of the mid-1970's. Early papers dealt with packetized voice transmission in the context of long-haul networks (see, for instance, [12, 13, 15, 16, 17]), and recognized the advantages of the packetized voice technique. With the increase in popularity of local area networks, focus shifted to integrated voice/data transmission on LANs. Most of this research has examined the performance of CSMA/CD networks with voice and data users [18, 19, 20, 21, 22, 23], but a small number of papers exploring Fasnet [6], token bus [24], register insertion [10], and token ring [11] networks have also appeared.

Packetized voice transmission on CSMA/CD networks was first studied by Johnson and O'Leary [19]. They examined the performance of a CSMA/CD network with digital voice terminals and no data users. Through computer simulations, the feasibility of several protocols with distributed control was verified, and through analytic techniques the performance of each protocol was predicted. Their analytic results, while ignoring the voice packet correlation issue, indicated that one packet speech buffering at source users provides sufficient support for many speech terminals with differing characteristics. Johnson and O'Leary observe that 82 speech terminals with voice coding rates of 16 Kbit/s can be supported by a 1 Mbit/s Ethernet with an average queueing delay of approximately 30 ms, and a voice packet size of 20 ms.

The performance of a 10 Mbit/s Ethernet under a combined voice and data load was simulated by Nutt and Bayer [20]. Results showed that a maximum of 93 simultaneous voice conversations could be supported with 5 ms voice packets and a voice encoding rate of 64 Kbits/s. With a data load of less than 5 percent and 50 voice conversations on the network, fewer than 0.5 percent of the voice packets experienced delays longer than the packet generation period. Nutt and Bayer also explored the effectiveness of various collision backoff algorithms. Traffic intensity and voice packet correlation were shown to have an important effect on network performance.

Various modifications of the CSMA/CD protocol to support voice and data transmission have also been proposed. Maxemchuk [21] presented a new protocol which allows any mix of voice and data traffic, and guarantees an upper bound on the delay of periodic (e.g., voice) traffic. In the protocol, data users use the standard CSMA/CD protocol, while voice users use a subset of CSMA/CD. Voice users monitor the bus to determine if the channel is idle before transmission. Once a voice user acquires the channel, he transmits all of his accumulated voice data. Consequently, voice users appear to have a dedicated time-division multiplexed slot. Through simulations of a 3 Mbit/s CSMA/CD network, the modified protocol was shown to reduce collisions, thus decreasing the network delay of the standard CSMA/CD protocol.

Combined voice and data transmission on a unidirectional bus system called Fasnet was discussed by Limb and Flamm [6]. They described an algorithm designed to facilitate the special features of both voice and data traffic. The protocol allocates a virtual channel for the duration of a talkspurt, and relinquishes the channel during silence intervals. Unused channel capacity can be utilized by data traffic. Limb and Flamm simulate a 20 Mbit/s Fasnet system with speech packets of 10 ms and PCM encoding of 64 Kbits/s. They observe that approximately 100 voice stations may share 50 virtual circuits with less than 2 percent of the talk spurts clipped by more than 50 ms.

A comparison of voice and data transmission on CSMA/CD networks and token buses was made by DeTreville [24]. A simulation of a 10 Mbits/s Ethernet with 64 Kbits/s speech coding, a voice packet size of 5.75 ms, and less than a 1 percent voice packet loss, showed approximately 60 voice conversations could be supported when no data traffic was present. With 5 percent data loading, approximately 35 voice conversations could be supported. DeTreville also simulated a token bus network based on the same parameter values as Ethernet. Results showed the token bus to have significantly more capacity than the CSMA/CD network. The CSMA/CD network supported voice traffic well until it reached a point of complete collapse. On the other hand, the token bus's performance degraded gradually without catastrophic failure. DeTreville concluded the token bus is superior to CSMA/CD for voice traffic.

Listanti *et al.* [10] examined the performance of ring and bus LANs under voice traffic. Simulation studies were based on 1 Mbit/s networks. The voice coding rate was assumed to be 32 Kbits/s. Results showed that, for a variety of packet lengths, the ring protocols simulated (i.e., register insertion, token passing, and slotted) have lower average delays than the CSMA/CD network. Listanti *et al.* observed the lowest average delays were achieved by the register insertion protocol.

This paper discusses the performance of a token passing ring network with packetized voice and data traffic (see [11] for a preliminary study by the author of this paper). As mentioned in the introduction, little has been established about the performance of a token passing ring network when it is subjected to a voice load in addition to a data load. Past research has focused on analyzing its performance for data applications. Hence, this paper evaluates the performance of a token passing ring network through extensive simulations to determine how well the token passing ring network suits real time voice and data integrated communications. This paper also explores several interesting issues which have not been addressed in the previous research on voice and data integrated transmission. The changes in network performance that occur when data traffic intensity is widely varied are discussed, as are the effects of voice and data packet lengths on network performance.

3. Simulation Model

3.1 Network Model

In our simulation model, two types of users are assumed on a token passing ring network: voice and data users. Fig.1 demonstrates the connection of voice users to our token passing ring network. At each voice user, a continuous voice analog signal is digitized by a coder. For instance, a typical PCM encoder operating at 8 KHz produces one 8-bit word every 125 μsec. The generated samples are accumulated in a packetizer. When the number of samples in the packetizer reaches the predetermined packet length, a header is attached and a voice packet is generated. The generated voice packet is then examined by a speech activity detector to see if it contains some minimal level of speech activity. Silent packets are discarded. Nonsilent packets are stored in the buffer in the order of their generation, and await transmission. The packet generation cycle is independent of the packet transmission process; thus, the queue size at the buffer continues to grow while a packet is waiting for transmission. Since we assume infinite buffer capacity, there is no packet loss at the voice source user.

Transmission control is based on a token, which is passed from user to user around the ring in a predetermined sequence. No priority is assumed for voice users over data users; both types of users are treated equally. When a free token is passed to a user ready to send a packet, the user changes the token status to busy and appends a packet to the token. The packet circulates around the ring to the intended destination user and then returns to the transmitting user, thereby acknowledging successful transmission of the packet around the ring. The token is then released downstream to the next user with packets available for transmission.

In general, we assume that only the head packet of a user buffer can be served by a token (i.e., limited service). The packets remaining in the buffer must wait for later visits of the free token. To minimize the loss of voice packets due to excessive transmission delays, the variance in transmission delays should be kept small. Intuitively, it is clear that the limited service discipline gives a small variance in delays. To verify this intuition, we compare the limited and exhaustive (i.e., where the token serves a user until his buffer is emptied) service disciplines in Sec.5.3(D).

No specific play-out scheme is assumed. Voice packets are played out immediately upon their arrival at a destination user. Infinite buffer space for storing incoming voice packets is assumed at destination users, and hence, there is no packet loss at the destination due to buffer overflow.

3.2 Data and Voice Traffic Models

In our simulations, both voice and data traffic are modeled. Data packet arrivals at a user are modeled by a Poisson process, and data arrivals at different users are independent of each other. The traffic generated by a Poisson process may not be as bursty as real data traffic, but the difference is expected to be relatively unimportant in determining the effect of data traffic upon voice traffic.

The termination and generation of voice calls is not considered in our simulations. All voice users are assumed to be active with calls throughout the entire simulation period. This assumption is made because statistical fluctuations in the presence of talkers are much slower than statistical fluctuations in the generation and transmission of voice and data packets. The holding time per voice call is on the order of hundreds of seconds, while the holding times of voice and data packets are on the order of tens of milliseconds.

Since we assume consistently active voice users, voice traffic at a user is modeled as having alternating talkspurts and silences (Fig.2). When a call is in talkspurt mode, voice packets of constant length are generated periodically in W_p time intervals, where W_p is the voice packet generation period (i.e., the time required to accumulate enough voice samples to construct a packet). If we let

V : voice coding rate (bits/sec),

P_v : voice packet length excluding header (bits), and

H : voice packet header length (bits),

W_p can be defined as:

$$W_p = \frac{P_v}{V} \qquad (3-1)$$

where a generated voice packet has $P_v + H$ bits in total. During silences, no packets are generated.

The statistics for the duration of talkspurts and silences in conversational speech depend on the mechanism used to detect speech activity. For instance, the average lengths of talkspurts and silences are assumed to be 1.23 and 1.77 seconds in [14], 0.185 and 1.31 seconds in

[21], and 0.17 and 0.41 seconds in [25], respectively. The effects of various talkspurt/silence statistics on network performance are also examined in this paper.

4. Performance Measures

Since excessive delays can have serious disruptive effects on human conversation, voice packets must be received at the destination user within a fixed amount of time after their generation at the source user. Those packets that do not arrive within this time bound are considered lost and are discarded upon their arrival at the destination user. A small number of lost packets has been shown to have little, if any, effect on human speech intelligibility.

If the transmission delay (D_v) of a voice packet is defined as the time interval between the beginning of its digitization and the time it is played out at its destination, then D_v becomes

$$D_v = W_p + W_q^v + T_v + R_{o,d}. \qquad (4-1)$$

W_p is the voice packet generation period given by Eq.(3-1). Note, this period depends on the packet length chosen. Further delay (i.e., the voice queueing delay (W_q^v) at the source user) is incurred while gaining access to the network, and in the actual transmission of the packet (T_v). T_v is given by

$$T_v = \frac{(P_v + H)}{C} \qquad (4-2)$$

where C is the channel speed (bits/sec). $R_{o,d}$ (sec) is the sum of the bit latency and propagation delay between the origin and destination users. Bit latency is the delay introduced at users to monitor and change the token bit pattern. 1-bit latency is assumed at each user throughout our simulations.

Since packets which do not arrive within a specified control time (τ) are considered lost, the loss probability of voice packets is

$$P_{loss} = Prob[D_v > \tau] = 1 - Prob[D_v \le \tau] = 1 - D_v(\tau) \qquad (4-3)$$

where $D_v(x)$ is the distribution function of the voice packet transmission delay (D_v).[2]

The traffic intensity for voice users on the network (ρ_v) can be calculated by the following equation:

$$\rho_v = N_v \times \left(\frac{\frac{\overline{T}}{(\overline{T}+\overline{S})} \times V}{P_v} \right) \times \left(\frac{P_v + H}{C} + R_{o,d} \right) \qquad (4-4)$$

where N_v is the number of voice users on the network, $\overline{T}$ and $\overline{S}$ are the average lengths of talkspurts and silences, respectively, and V is the voice encoding rate. The second factor of

2 Some method of time-stamping must be used during voice packet generation, since destination users need to determine when the transmission delay of a voice packet exceeds the control time (τ) and, thus, the packet should be discarded.

this product represents the number of packets generated by a voice user of the network. The third factor represents the amount of time a voice packet occupies the channel.

The transmission delay D_d of a data packet is defined in a similar manner:

$$D_d = W_q^d + T_d + R_{o,d} \qquad (4-5)$$

where W_q^d is the queueing delay of a data packet at the source user, and T_d is the transmission time of a data packet. T_d is given by $T_d = P_d/C$, where P_d is the data packet length (including header) in bits.

Data traffic intensity on the network is defined as follows:

$$\rho_d = N_d \lambda \frac{P_d}{C} \qquad (4-6)$$

where N_d is the number of data users, and λ is the arrival rate of data packets at a data user.

5. Simulation Results

5.1. Parameter Values

To explore the effectiveness of a token passing ring local area network with packetized voice and data traffic, extensive simulations were performed. In the simulation model used, the length of the network ring is assumed to be 1 km, and voice and data users are distributed uniformly around the ring. All arrivals are independent of each other, and data users have identical Poisson arrival rates. The values used for the token length (3 octets), the packet header length H (21 octets), and the channel speed C (1 Mbits/sec) are based on the IEEE standard 802.5 for token ring local area networks [5]. We assume 1-bit latency at each user, and a propagation delay of 5 μsec per km of cable. It should be noted, that selecting this value for the propagation delay represents an optimistic point of view, since other delay components (e.g., signal rise times or repeater delays) may significantly increase the total end-to-end propagation delay [8].

Throughout our simulations, the distributions of talkspurts and silences are exponential. With the exception of the simulation data in Fig.8, the average lengths of the talkspurts ($\overline{T}$) and silences ($\overline{S}$) within calls are assumed to be 0.17 and 0.41 seconds, respectively [25]. The PCM voice coding scheme is assumed at all voice users. This scheme allows the digitization of voice data at a rate of 64 Kbits/sec. We assume voice users are active throughout the entire simulation period, and, hence, N_v is actually the number of co-existing voice calls on the network. A limited service discipline is assumed at all users except those in Figs.16, 17, and 18 where an exhaustive service discipline is used.

5.2. Numerical Examples for a Network with Voice Users Only

This section describes the simulation results for a token passing ring network used exclusively by voice users.

A. The Effects of Voice Packet Length on Performance

Fig.3 displays the average transmission delay ($\overline{D_v}$) of voice packets as a function of the voice packet length (P_v). The transmission delay components—W_p (the voice packet generation period), $\overline{W_q^v}$ (the average queueing delay), and T_v (the transmission time of a voice packet)—are also shown. The relationship of these components can be seen in Eq.(4-1). The values for $\overline{D_v}$ and $\overline{W_q^v}$ were obtained through simulations, while the values for W_p and T_v were calculated from Eqs.(3-1) and (4-2), respectively. Note, compared with the values of the other transmission delay components, the value of the bit latency and propagation delay ($R_{o,d}$) is negligible (i.e., 35 μsec), and is not shown. In this figure, 30 voice users are active on the network.

In Fig.3, the optimal voice packet length (P_{opt}) to minimize the transmission delay ($\overline{D_v}$) is 640 bits. The P_{opt} is determined by the following trade-off: when the voice packet length is increased, the packet generation period (W_p) becomes larger, adding to the queueing delay of the packet; on the other hand, fewer voice packets are generated during a talkspurt, causing less packet header overhead, and thus reducing the transmission delay of the voice packet. The voice traffic intensity (ρ_v) can be calculated from Eq.(4-4). Solving this equation for $\rho_v = 1$ shows that the theoretical saturation point of the network occurs at P_v equal to 248 bits. In our simulation model, the value of P_v at 248 bits causes a transmission delay of 1.471 seconds, and the value of P_v at 240 bits causes an infinite delay. This close correspondence between the theoretical and simulation saturation points is a good indication that the simulation model being used is valid.

The probability density functions for the queueing delay of voice packets (W_q^v) are shown in Fig.4. When voice packet lengths (P_v) are equal to 320, 640, and 1920 bits, the network voice traffic intensities (ρ_v) equal 0.89, 0.72, and 0.61, respectively. Fig.5 shows the packet loss probability (P_{loss}) as a function of the voice packet length (P_v). The control time (τ) is used as the parameter to this figure, and voice packets whose transmission delays exceed τ = 40, 50, 60, 100, and 150 msec are considered lost. The number of voice calls (N_v) is 30 in both figures. When τ is equal to 50 msec, the voice packet length must be (approximately) within the range $880 \leq P_v \leq 1840$ bits to guarantee $P_{loss} \leq 1\%$.

B. Maximum Number of Voice Calls Supported

Fig.6 shows the average delay of voice packets ($\overline{D_v}$) for various values of P_v as a function of the number of voice users (N_v). Three loss probability (P_{loss}) curves (i.e., τ = 50, 100, and 150 msec) for P_v equal to 1280 bits are also given in the figure. The sections of the voice delay curve (P_v = 1280 bits) above the 1% loss range for τ = 100 and 150 msec show the network operating in an unstable situation (i.e., in a range where the delay curve is very steep). In this unstable situation, a small fluctuation in the number of voice users can easily cause the network to go beyond its saturation point. Because of this unfavorable characteristic, we will use τ equal to 50 msec for the remainder of this paper. It is also worth noting that in an internetworking environment smaller control times (τ) are preferred. When networks are interconnected, end-to-end transmission of packets may involve several networks; thus, minimizing the control times for individual networks may be necessary to satisfy the end-to-end transmission delay constraints essential to voice users [10].

Theoretical saturation points can also be calculated for the curves in Fig.6. For $P_v =$ 640 bits, saturation occurs at 42 active voice users. For $P_v = 1280$ bits, saturation occurs at 47 active voice users. Through our simulations, the values 41 and 46 corresponding to $P_v = 640$ and 1280 bits, respectively, were obtained when the network reached its saturation point. Again, these simulation values compare favorably to the theoretical saturation values calculated. Note, the delay curves for $P_v = 640$ and 1280 cross in this figure. Their intersection is caused by the effect of header bits with different packet lengths. Reduced voice packet lengths cause more packets to be generated and, thus, more packet header bits must be transmitted on the network. The extra overhead resulting from these header bits causes the network to reach its saturation point sooner than a network with a larger voice packet length.

Fig.7 shows the maximum number of voice users (N_v^{max}) allowed on a network for various voice packet lengths (P_v) while satisfying the constraint $P_{loss} \leq 1\%$. Curves for values of τ equal to 40, 50, 60 and 100 msec are included. For instance, if $\tau = 50$ msec, a voice packet of length $880 \leq P_v \leq 1800$ bits achieves the maximum value of $N_v^{max} = 30$. Considering that voice users are, in reality, not always active, a network supporting 30 concurrent conversations may be quite sufficient. In fact, often the number of voice communication lines connected to a network is 5–10 times the number of voice users that can be supported by the system at a single time [20]. Thus, our network could support 150 to 300 voice lines, satisfying the IEEE 802 token ring standard of supporting a maximum of 250 users on a network [5].

C. Optimal Voice Packet Length

Tab.1 shows the loss probability (P_{loss}) and the average delay of voice packets ($\overline{D_v}$) for the case when $\tau = 50$ msec. Information on voice packet lengths (P_v) is given in both bits and msec. The values in this table are taken from points along the curve for $\tau = 50$ msec in Fig.7. In this table, all eight values of P_v satisfy $P_{loss} \leq 1\%$. Of the eight, three values of P_v (i.e., 960, 1280, and 1600 bits) achieve the maximum value of 30 voice users, but the average transmission delays of the three values vary by over 10 msec. In this situation, the shorter the voice packet length, the smaller the average network delay. This is due to the influence of the packetization delay (W_p) on $\overline{D_v}$. As shown in Fig.3, W_p has a significant effect on $\overline{D_v}$ for sufficiently large voice packet lengths. Tab.2 contains the loss probability (P_{loss}) and the average delay of voice packets ($\overline{D_v}$) for the case where $\tau = 50$ msec and the number of voice users is held constant at 28. Again, information on voice packet lengths is given in both bits and msec. The values for this table were obtained by slicing the curve for $\tau = 50$ msec from Fig.7 by the line $N_v = 28$. In this table, $P_v = 640$ gives the smallest delay while satisfying $P_{loss} \leq 1\%$. But the resulting loss probability is 0.74%, and such a large value of P_{loss} may not be desirable, since even small fluctuations in network traffic may cause the loss probability constraint to be violated.

The selection of a voice packet length (P_v) to maximize the number of voice users (N_v) on a network depends on the performance criteria used. Smaller voice packet lengths result in smaller average transmission delays ($\overline{D_v}$), as well as increased voice packet loss probabilities (Tabs.1 and 2). Other performance criteria, not discussed in this paper, may also be considered. For example, voice packet length can have an affect on the quality of voice conversations. When P_v becomes too large, loss of one packet may be significant, and successive loss of long voice packets may result in voice conversations of reduced quality.

D. The Effects of Talkspurt/Silence Length on Performance

Fig.8 shows N_v^{max} for various talkspurt/silence statistics. Voice packets whose delays exceed 50 msec are considered lost, and the loss probability constraint is enforced at less than or equal to 1%. The average lengths of talkspurts ($\overline{T}$) and silences ($\overline{S}$) are the parameters varied, and their distributions are assumed to be exponential. For each pair of ($\overline{T}$, $\overline{S}$), the ratio $\overline{T}$: $\overline{S}$ is kept constant and is equal to 0.17 : 0.41. N_v^{max} decreases as the average talkspurt length $\overline{T}$ becomes larger. This occurs because a longer talkspurt period gives a larger correlation in voice packet generation (i.e., produces a larger number of packets per talkspurt). Less correlation in arrivals implies that more users can be supported by the network.

5.3. Numerical Examples for a Network with Voice and Data Users

The following section discusses networks with both voice and data users. Voice users and data users are uniformly distributed on the ring. Poisson arrivals of fixed length packets are assumed at each data user. Data packet lengths of 1024 and 8192 bits (including header) are assumed in our simulations. 1024 bit data packets are typical of those found in current networks and represent short interactive data. 8192 bit data packets are used for bulk data transfer (i.e., file transfer, image transfer, etc.). In the IEEE 802 standard for token passing ring networks, users are allowed to hold a token (i.e., allowed to transmit) for up to 10 msec. Thus, on a 1 Mbit/sec channel, users may transmit a maximum packet of up to 10,000 bits. The packet length of 8192 bits used in our simulations is close to the situation where a user holds the token for the maximum allowable time period.

A. The Effects of Data Traffic Intensity on Performance

Fig.9 shows the packet loss probability (P_{loss}) as a function of the voice packet length (P_v) for various data traffic intensities (ρ_d). All voice packets whose delays exceed 50 msec are considered lost, and the data packet length (P_d) equals 1024 bits. The number of voice calls (N_v) and data users (N_d) on the network are identical and equal to 15. The network supports data traffic up to approximately $\rho_d = 0.45$ without violating the 1% loss constraint. This level of ρ_d seems quite good. For instance, measurements of the data load of a typical 10 Mbit CSMA/CD network have shown that the average network bandwidth data utilization is less than 5% [20,24,26], so token ring networks with voice and data traffic may effectively handle real-world network environments. Note the relationship between curves in this figure; linear differences in data traffic intensities result in exponential differences in values along the loss probability axis.

B. Maximum Number of Voice Calls Supported

Figs.10 and 11 show the average delays of voice packets ($\overline{D_v}$) and data packets ($\overline{D_d}$), respectively, as a function of the number of voice calls on a network. In both figures, N_d is 15, data traffic intensity (ρ_d) is equal to 0.3, and P_d is equal to 1024 bits. In Fig.10, theoretical calculations of network saturation points were performed and compared with the saturation points gathered through simulation. Saturation of the network occurs when $\rho_v + \rho_d = 1$, where ρ_v is given by Eq.(3-4) and ρ_d is given by Eq.(3-6). For example, the theoretical saturation points for $P_v = 960$ and 1600 bits were $N_v = 31$ and 33, respectively. The simulation saturation points for the identical values of P_v occurred at $N_v = 31$ and 32. In fact, the theoretical and simulation saturation points were very close (i.e., within two voice users) for all values of P_v in Fig.10.

In Fig.11, the data delay curves for a variety of values of P_v all lie close together, except near the network saturation points. Near these saturation points, the constraint $P_{loss} \leq 1\%$ is violated. For example, when $\tau = 50$ msec, the voice packet length is 2240 bits, and there are 18 voice users on the network, P_{loss} equals 1.14%. This suggests that data delays are not significantly affected by voice packet lengths in the range where real-time voice constraints are satisfied. On the other hand, as seen in Fig.10, voice delays are very sensitive to the value of P_v selected. Thus, it should be possible to adjust P_v to maximize the number of voice users, with minimal effect on the performance received by data users.

Fig.12 shows the maximum number of voice calls (N_v^{max}) allowed on a network without violating loss condition $P_{loss} \leq 1\%$. The parameter varied in this figure is the control time (τ). Again, N_d is 15, P_d is 1024 bits, and data traffic intensity (ρ_d) is equal to 0.3. For example, voice packets of length (approximately) $1280 \leq P_v \leq 1600$ bits achieve the maximum value of $N_v^{max} = 20$ when $\tau = 50$ msec. Note, for most values of P_v, the curves in Figs.7 and 12 vary by approximately 10 voice users. The network in Fig.7 has voice users only; thus, this difference in the maximum number of voice calls supported is caused by the data traffic intensity of 0.3 carried by the network in Fig.12.

Fig.13 shows the number of voice calls (N_v) allowed on a network for different data traffic intensities (ρ_d). In the figure, the value of $P_d = 1024$ bits and the value of $\tau = 50$ msec. When $\rho_d = 0$, the maximum number of voice calls (N_v^{max}) is 30. This maximum can be achieved for P_v in the range $880 \leq P_v \leq 1800$ bits. When $\rho_d = 0.3$ the maximum number of voice calls is 20. Note, unlike the curve for $\rho_d = 0$, this curve for $\rho_d = 0.3$ does not have a long horizontal segment. Thus, to achieve a maximum value for N_v^{max}, the network is highly sensitive to the voice packet length (P_v) selected.

C. The Effects of Data Packet Length on Performance

The preceding discussions were all based on fixed length data packets of 1024 bits. In this subsection, we assume a data packet length of $P_d = 8192$ bits and examine the network performance changes that occur. Fig.14 shows the voice packet loss probability (P_{loss}) as a function of the voice packet length (P_v) for various data traffic intensities (ρ_d). In this figure, the data packet length (P_d) is 8192 bits and τ is 50 msec. There are 15 voice and 15 data users on the network. The increase in the data packet length reduces the values of ρ_d that satisfy the constraint $P_{loss} \leq 1\%$. In Fig.9, with $P_d = 1024$ bits, the network supports data traffic up to approximately $\rho_d = 0.45$ without violating the 1% loss constraint. However, the network in Fig.14 only supports data traffic up to approximately $\rho_d = 0.25$ while maintaining $P_{loss} \leq 1\%$. The results in sections 4.3(A) and (B) assume a data traffic intensity of 0.3 when $P_v = 1024$ bits. When $P_v = 8192$ bits, the network cannot support this intensity, so $\rho_d = 0.2$ is used in our simulations for Figs.15, 17, and 18. Also, note, that the ranges of voice packet lengths which satisfy the 1% loss constraint in Fig.14 are narrower than the ranges which satisfy the constraint in Fig.9.

Fig.15 compares the maximum number of users allowed on a network when the data packet lengths (P_d) are equal to 1024 and 8192 bits. In this figure, there are 15 data users, and the data traffic intensity (ρ_d) equals 0.2. For the values of τ in this figure (i.e., $\tau = 50$ and 100 msec), the curves for the larger data packet length fall below those for the smaller P_d. For example, when $\tau = 50$ msec, if the data packet length is 1024 bits, N_v^{max} equals 23, while N_v^{max} equals 18 when the data packet length is 8192 bits. Thus, in this case, a data packet length of 1024 bits provides better network performance.

D. Limited vs. Exhaustive Service

The curves in Figs.3 through 15 display data from simulations where a limited service discipline is employed. In these networks, a user may place only one packet on the network, before releasing the token. In this section, we compare limited and exhaustive service. In exhaustive service, a user may continue to place packets on the network until his transmit queue is empty. Then he releases the token. Fig.16 shows the voice packet loss probability (P_{loss}) of a network as a function of the voice packet length (P_v) in the case where an exhaustive service discipline is used. The data packet length for this figure is 8192 bits. The values of the data traffic intensities (ρ_d) shown in Fig.16 have slightly higher loss probability curves than those shown in Fig.14. For example, the curve for $\rho_d = 0.25$ in Fig.14 satisfies the 1% loss probability constraint for some voice packet lengths, while in Fig.16, the curve for $\rho_d = 0.25$ cannot satisfy the same constraint for any voice packet length. Because exhaustive service allows users to hold the token for unspecified lengths of time, excessive transmission delays may result, causing loss of voice packets and increased P_{loss}. Performance of the limited service discipline results in lower values for P_{loss}, since it minimizes the variance in transmission delays.

Fig.17 shows the average data packet transmission delay ($\overline{D_d}$) as a function of the number of calls on the network. In this figure, 15 voice and 15 data users are active, and the data packet length (P_d) is 8192 bits. Statistics for both limited and exhaustive service are displayed. Note, for most values of N_v in the figure, there is no significant difference in the limited and exhaustive service cases. Discrepancies do exist where the number of calls on the network approaches the level of network saturation. Data packet delays rise more sharply in the exhaustive service case, but this rise occurs in an area where $P_{loss} > 1\%$. For instance, $P_{loss} \leq 1\%$ is violated when the voice packet length is 960 bits, 18 voice users are active on the network, and $\tau = 50$ msec. We are concerned with networks which meet the real-time constraints of speech, and, thus, the portions of the curves which satisfy $P_{loss} \leq 1\%$. This figure suggests that the service discipline used—limited or exhaustive—has little, if any, affect on the average data packet delay when $P_{loss} \leq 1\%$ is satisfied.

Fig.18 compares the number of voice users (N_v) allowed on a network when limited and exhaustive service disciplines are used. In this figure, there are 15 data users, the data packet length (P_d) is 8192 bits, the data traffic intensity (ρ_d) is 0.2, and the constraint $P_{loss} \leq 1\%$ is satisfied. For both control time values (i.e., $\tau = 50$ and 100 msec), the limited service discipline provides slightly better performance. For example, when $\tau = 50$ msec the maximum number of users supported by the network is 18 in the limited service case, and 17 in the exhaustive service case.

5.4 Summary of Simulation Results

Extensive simulations were performed to examine the effectiveness of a token passing ring local area network with both voice and data traffic. In the simulations, system parameters were manipulated, and their effects on network performance were examined. The following observations summarize the simulation results discussed in Secs.5.2 and 5.3.

(1) Voice packet length can have a significant effect on network performance. Selection of an optimal voice packet length is dependent on the performance criteria used. For example, when maximizing the number of voice users and minimizing the average voice delay of a voice user network with a 50 msec control time and a voice packet loss probability below 1%, our simulation results yielded a maximum of 30 voice users and an optimal voice packet length of 960 bits. Other criteria, such as assessing the effects of voice packet length on the quality of voice conversations, may also be considered in voice packet length selection.

(2) Smaller average talkspurt/silence lengths reduce the correlation of voice packet arrivals, and, thus, allow more users to be supported by the system.

(3) The data traffic intensity and control time (i.e., the time limit for determining which voice packets are lost) required for acceptable network performance in real-world environments can be satisfied by the token passing ring local area network.

(4) The data packet length can affect the voice performance of the network. In our simulations, the data packet length of 1024 bits out-performed the data packet length of 8192 bits (i.e., the extreme case where users try to hold a token for very close to the maximum time allowed in the IEEE 802 standard).

(5) A network with a limited service discipline yields slightly better performance than a network with an exhaustive service discipline in the range where the loss probability is less than or equal to 1%.

These simulation results suggest that token passing ring networks can effectively handle voice and data traffic in real-world network environments.

6. Conclusions

In this paper, simulation models were developed to evaluate the performance of packetized-voice/data transmission on a token passing ring network. The network performance measures obtained include: the distribution of transmission delays for voice packets, the average transmission delay and loss probabilities for voice packets, the number of voice users allowed on a network while satisfying the real-time constraints of speech, and the average transmission delay for data packets.

Simulation results suggest that system parameters (e.g., the voice packet length, data packet length, control time for voice packets) can be adjusted to allow token passing ring networks to support both voice and data traffic with acceptable performance. Using the data obtained through simulations, the following issues were examined: selecting the optimum voice packet length, the maximum number of voice calls supported by a network, the effects of talkspurt/silence lengths and data traffic intensity on network performance, and limited versus exhaustive service disciplines.

Due to the time-intensive nature of the simulations performed for this paper, confidence intervals were not calculated for the points in our figures.

Other topics related to voice and data traffic on token passing ring local area networks include: the effects of voice packet length on the quality of voice conversation, a performance comparison of token passing ring local area network with other local area networks, and, of course, a theoretical analysis of a token passing ring local area network with voice and data traffic. These issues await further research.

References

[1] W. D. Farmer and E. E. Newhall, "An Experimental Distributed Switching System to Handle Bursty Computer Traffic," Proc. of ACM Symp. Problems Optimization Data Comm. Systems, 1969.

[2] J. R. Pierce, "Network for Block Switches of Data," The BSTJ, July/August, 1972.

[3] E. R. Hafner, Z. Nenadal and M. Tschanz, "A Digital Loop Communications System," IEEE Trans. on Commun., Vol.COM-22, No.6, June, 1974.

[4] R. M. Metcalfe and D. R. Boggs, "Ethernet: Distributed Packet Switching for Local Computer Networks," Communications of the ACM, Vol.19, No.7, July, 1976.

[5] "Local Area Network Standards: Token-Passing Bus Access Methods and Physical Layer Specifications," IEEE Standard 802.4-1985, 1985.

[6] J. O. Limb and L. E. Flamm, "A Distributed Local Area Network Packet Protocol for Combined Voice and Data Transmission," IEEE Journal on Selected Areas in Communications, Vol.SAC-1, No.5, Nov., 1983.

[7] F. A. Tobagi, F. Borgonovo and L. Fratta, "Expressnet: A High Performance Integrated Service Local Area Network," IEEE Journal on Selected Areas in Communications, Vol.SAC-1, No.5, Nov., 1983.

[8] W. Bux, "Local Area Subnetworks: A Performance Comparison," IEEE Trans. on Commun., Vol.COM-29, Oct., 1981.

[9] B. W. Stuck, "Calculating the Maximum Mean Data Rate in Local Area Networks," IEEE Computer, May, 1983.

[10] M. Listanti, A. Pattavina, A. Roveri and F. Villani, "Evaluation of Multiaccess Protocols for Packet Voice Communication in Local Area Networks," Proc. of the ICC, 1985.

[11] T. Suda, C. Yuen and K. Ohtsuki, "Performance Evaluation of Packetized Voice Transmission on a Token Passing Ring Network," GLOBECOM 1985.

[12] D. Cohen, "Packet Communication of Online Speech," Proc. of the NCC, 1981.

[13] B. Gold, "Digital Speech Network," Proc. of the IEEE, Vol.65, 1977.

[14] P. T. Brady, "A Technique for Investigating On-Off Patterns of Speech," The BSTJ, Jan., 1965.

[15] J. G. Gruber, "Delay Related Issues in Integrated Voice and Data Networks," IEEE Trans. on Commun., Vol.COM-29, No.6, June, 1981.

[16] J. F. Kurose, M. Schwartz and Y. Yemini, "Multiple-Access Protocols and Time-Constrained Communication," ACM Computing Surveys, Vol.16, No.1, March, 1984.

[17] T. Bially, A. J. McLaughlin and C. J. Weinstein, "Voice Communication in Integrated Digital Voice and Data Networks," IEEE Trans. on Commun., Vol.COM-28, No.9, Sept., 1980.

[18] J. F. Shoch, "Carrying Voice Traffic through an Ethernet Local Network: A General

Overview," Local Networks for Computer Communications. A. West and P. Johnson (eds.), North-Holland, 1981.

[19] D. Johnson and G. O'Leary, "A Local Access Network for Packetized Digital Voice Communication," IEEE Trans. on Commun., Vol.COM-29, No.5, May, 1981.

[20] G. J. Nutt and D. L. Bayer, "Performance of CSMA/CD Networks Under Combined Voice and Data Loads," IEEE Trans. on Commun., Vol.COM-30, No.1. Jan., 1982.

[21] N. F. Maxemchuk, "A Variation on CSMA/CD That Yields Movable TDM Slots in Integrated voice/Data Local Networks," The BSTJ, Vol.61, No.7, Sept., 1982.

[22] P. Mui and N. D. Georganas, "Packet-voice Performance Evaluation of two Multiple Access Protocols with the use of an 8086-Based Real-time Simulator," Links for the Future, Science, Systems and Services for Communications, P. Dewilde and C.A. May (eds.), IEEE/Elsevier Science Publishers B.V. North Holland, 1984.

[23] M. Rios and N. D. Georganas, "A Hybrid Multiple-Access Protocol for Data and Voice-Packet Over Local Area Networks," IEEE Transactions on Computers, Vol. c-34, No. 1, Jan., 1985.

[24] J. D. DeTreville, "A Simulation-Based Comparison of Voice Transmission on CSMA/CD Networks and on Token Buses," The BSTJ, Vol.63, No.1, Jan., 1984.

[25] J. G. Gruber, "A Comparison of Measured and Calculated Speech Temporal Parameters Relevant to Speech Activity Detection," IEEE Trans. on Commun., Vol.COM-30, April, 1982.

[26] J. F. Shoch, J. A. Hupp, "Measured Performance of an Ethernet Local Network," Commun. Ass. Comput. Mach., vol. 23, Dec., 1980.

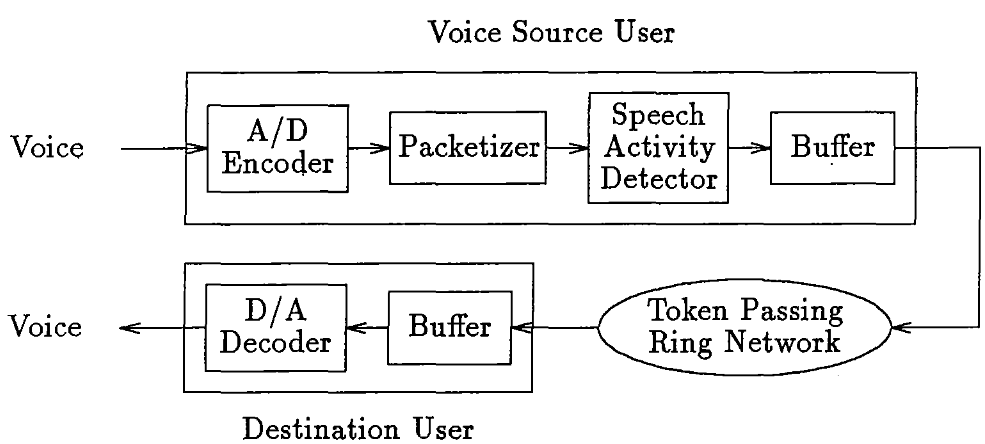

Fig.1 Voice Users

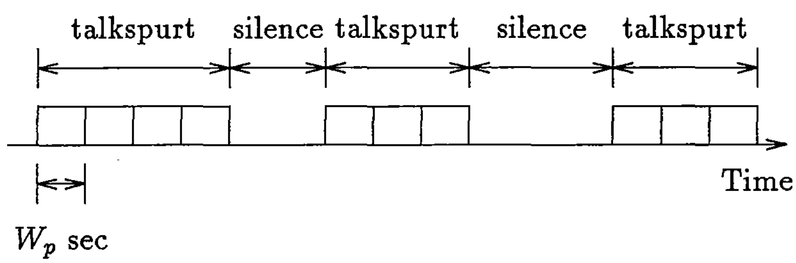

Fig.2 Correlated Arrival Process of Voice Packets

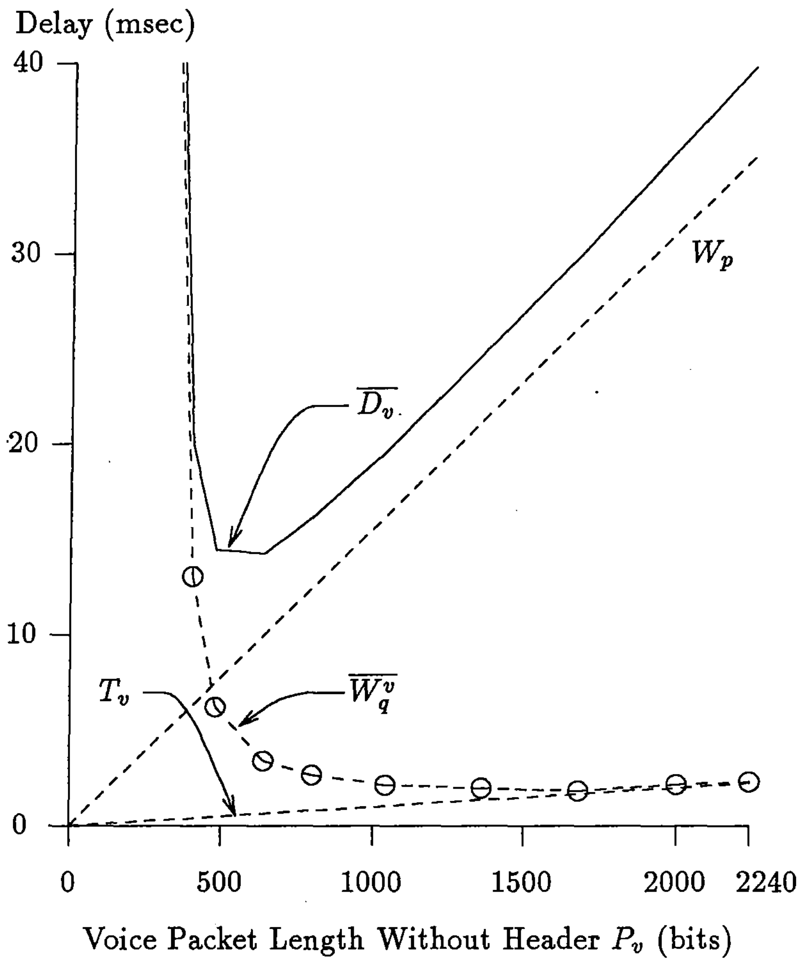

Fig.3 Average Transmission Delays (N_v = 30)

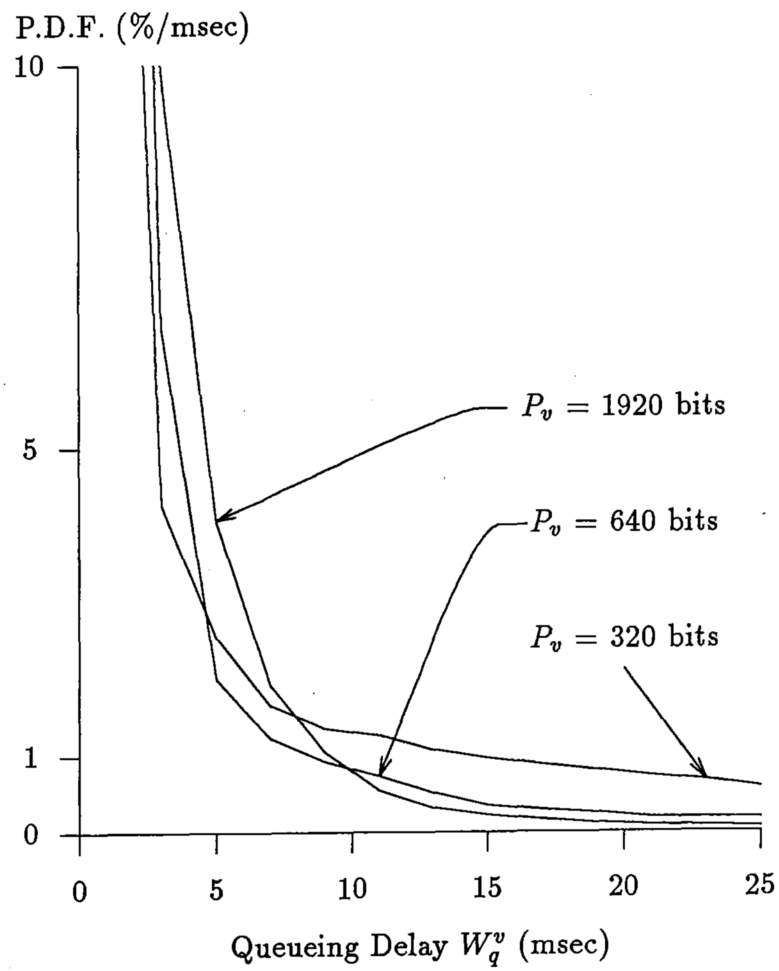

Fig.4 Probability Density Function of W_q^v ($N_v = 30$)

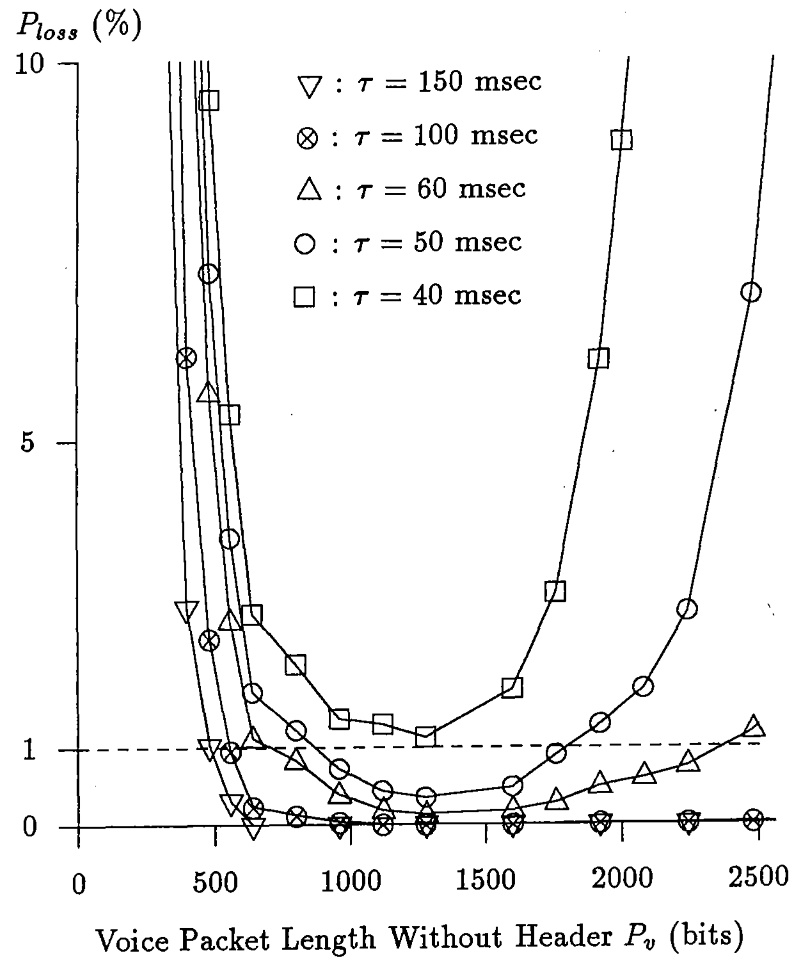

Fig.5 Loss Probability P_{loss} (N_v = 30)

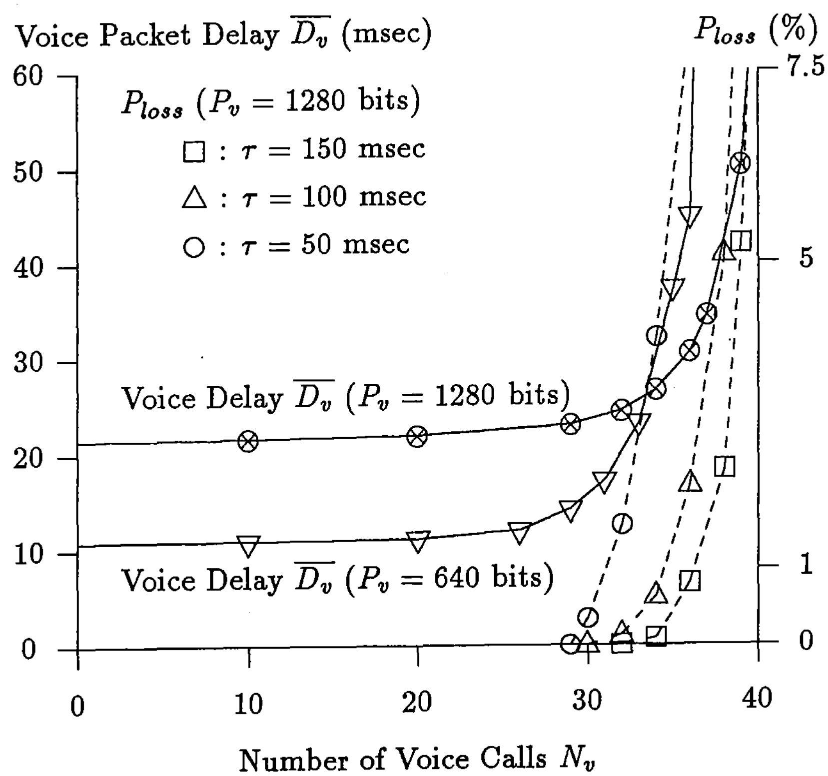

Fig.6 Transmission Delay and Loss Probabilities
(N_v = 30, P_v = 1280 bits)

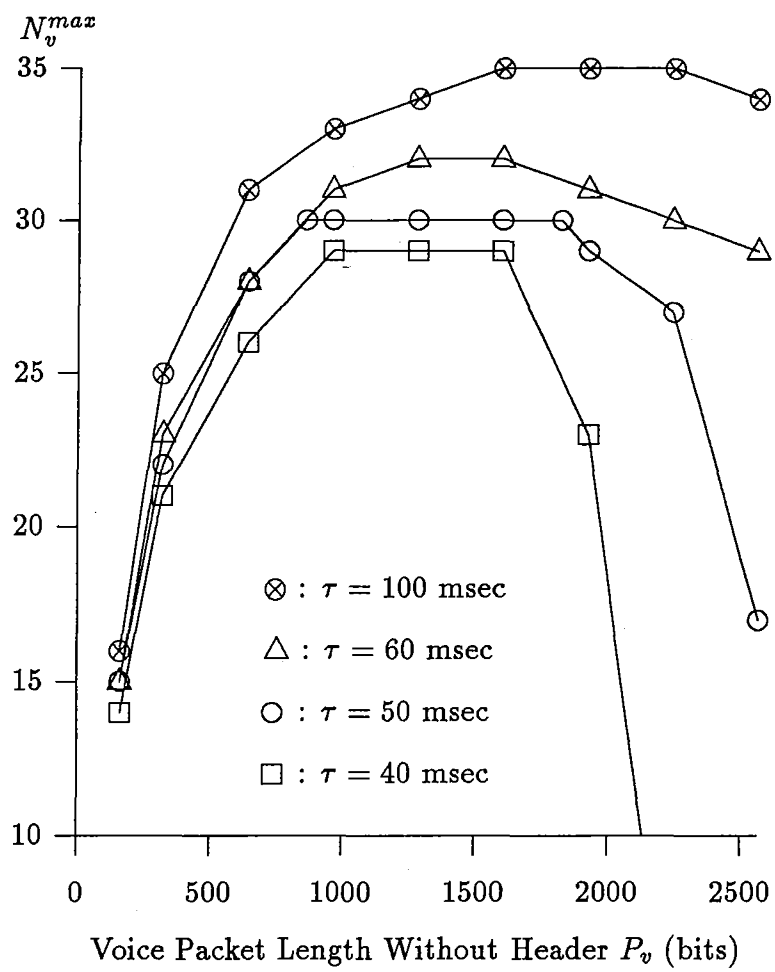

Fig.7 Maximum Number of Voice Calls Allowed on a Network

P_v	N_v^{max}	P_{loss} (%)	$\overline{D_v}$ (msec)
320 bits = 5 msec	22	1.00	7.55
640 bits = 10 msec	28	0.74	13.12
960 bits = 15 msec	30	0.72	18.27
1280 bits = 20 msec	30	0.35	23.31
1600 bits = 25 msec	30	0.48	28.68
1920 bits = 30 msec	29	0.61	33.97
2240 bits = 35 msec	27	0.76	39.02
2560 bits = 40 msec	17	0.94	43.50

Tab.1 Various Statistics ($\tau = 50$ msec)

P_v	N_v	P_{loss} (%)	$\overline{D_v}$ (msec)
320 bits = 5 msec	28	12.76	24.24
640 bits = 10 msec	28	0.74	13.12
960 bits = 15 msec	28	0.31	17.62
1280 bits = 20 msec	28	0.22	22.80
1600 bits = 25 msec	28	0.32	28.21
1920 bits = 30 msec	28	0.38	33.68
2240 bits = 35 msec	28	1.25	39.17
2560 bits = 40 msec	28	6.87	44.69

Tab.2 Various Statistics ($\tau = 50$ msec, $N_v = 28$)

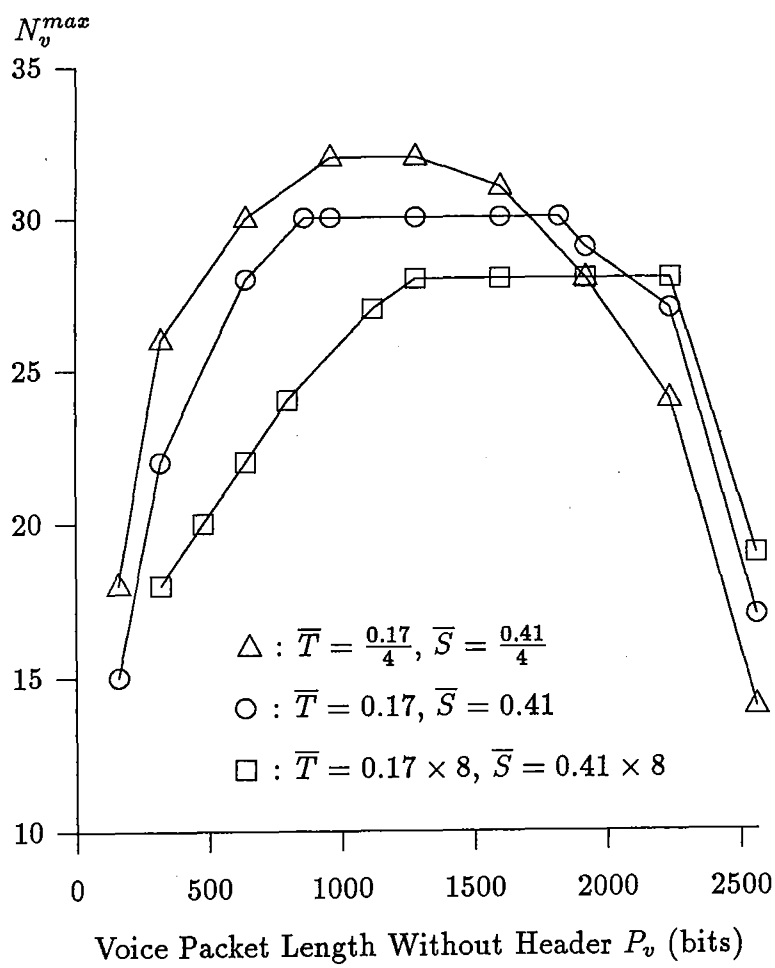

Fig.8 Maximum Number of Voice Calls Allowed on a Network ($N_v = 30$, $\tau = 50$ msec)

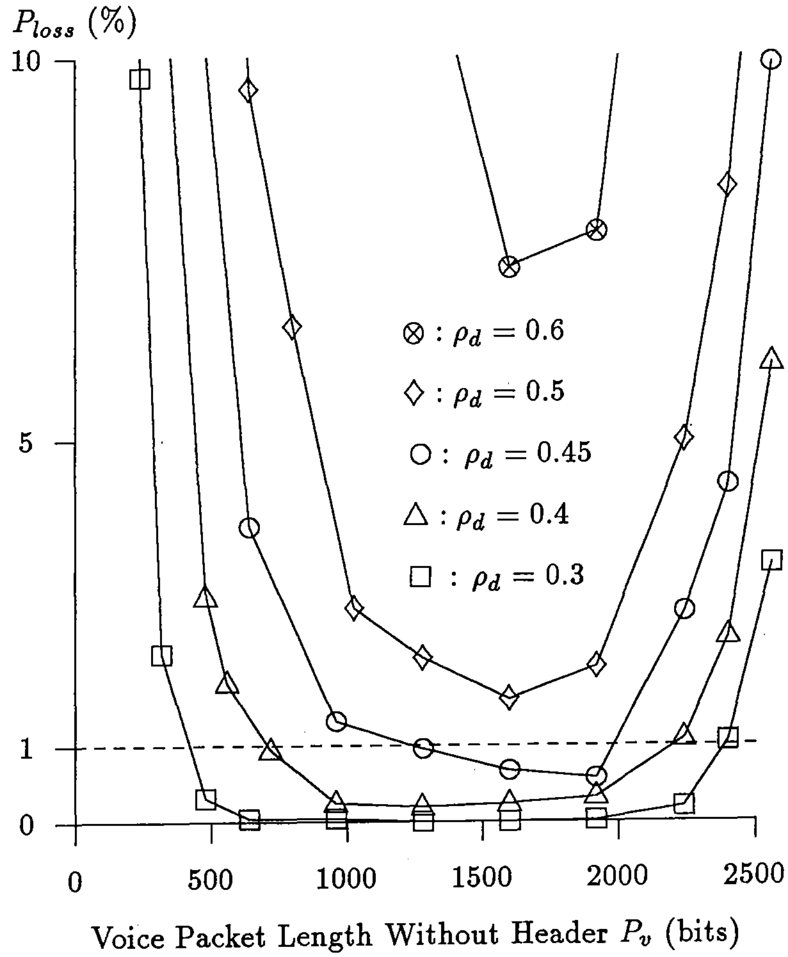

Fig.9 Loss Probability P_{loss}
($N_v = N_d = 15$, $P_d = 1024$ bits, $\tau = 50$ msec)

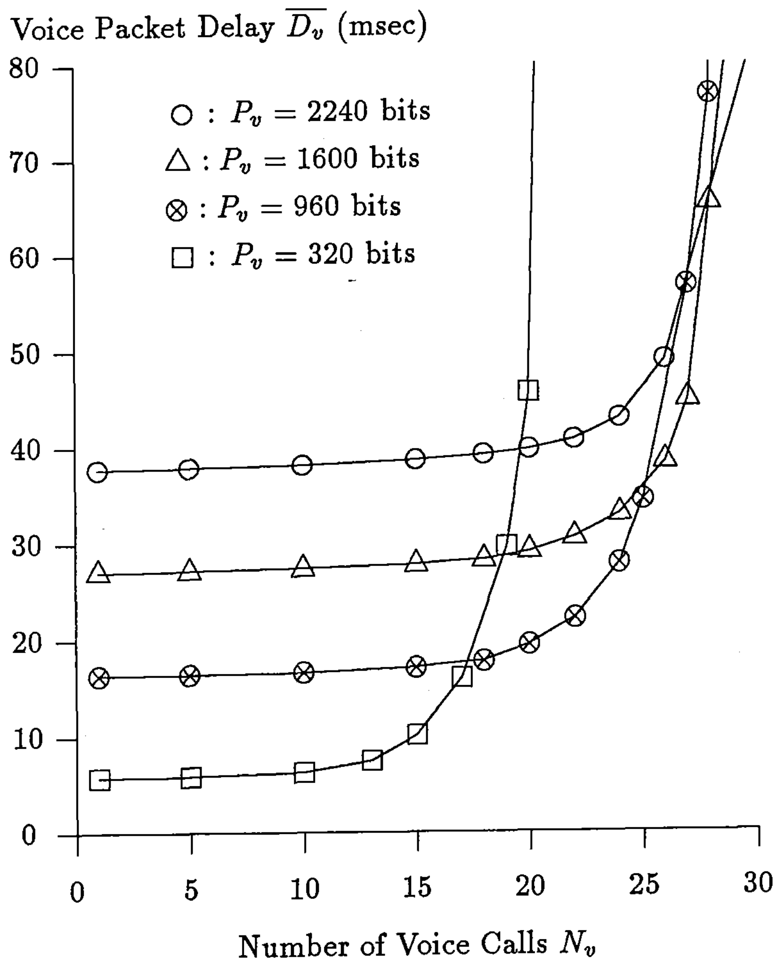

Fig.10 Voice Packet Transmission Delay $\overline{D_v}$
($N_d = 15$, $P_d = 1024$ bits, $\rho_d = 0.3$)

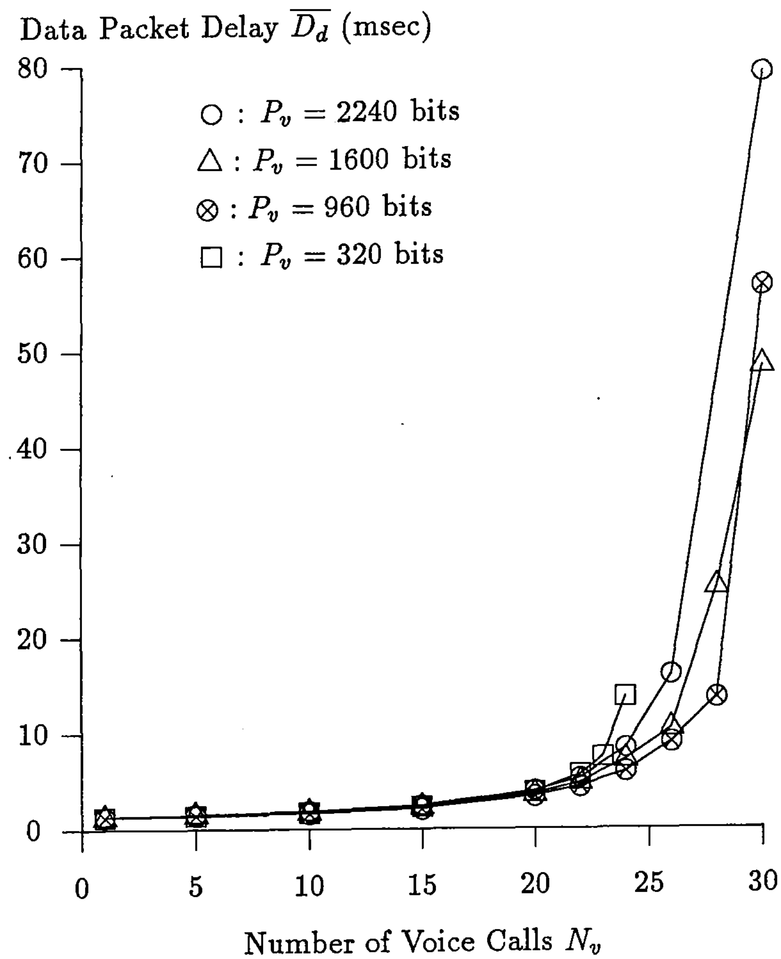

Fig.11 Data Packet Transmission Delay $\overline{D_d}$
($N_d = 15$, $P_d = 1024$ bits, $\rho_d = 0.3$)

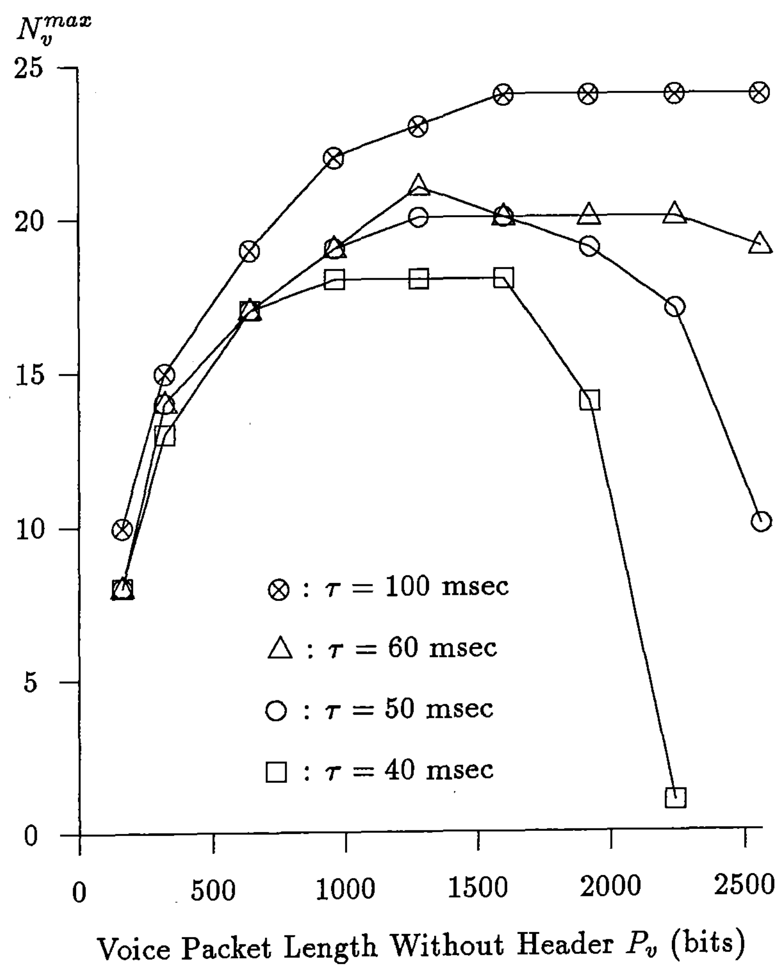

Fig.12 Maximum Number of Voice Calls Allowed on a Network
($N_d = 15$, $P_d = 1024$ bits, $\rho_d = 0.3$)

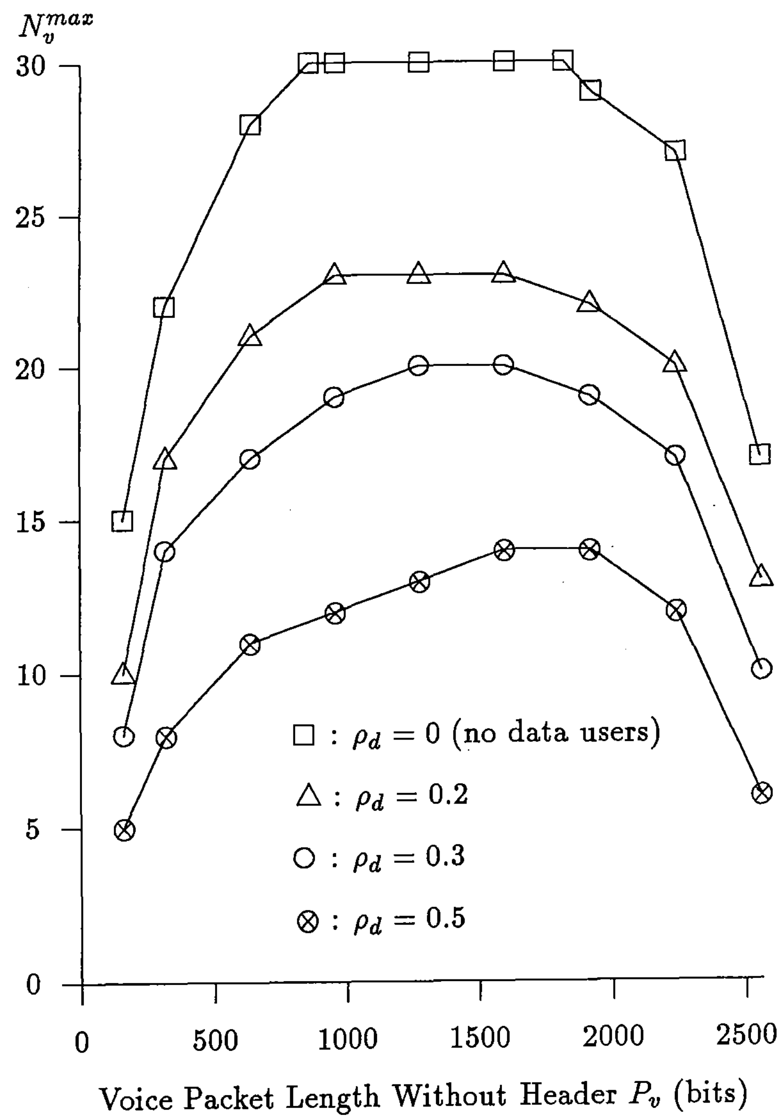

Fig.13 Maximum Number of Voice Calls Allowed on a Network
($N_d = 15$, $P_d = 1024$ bits)

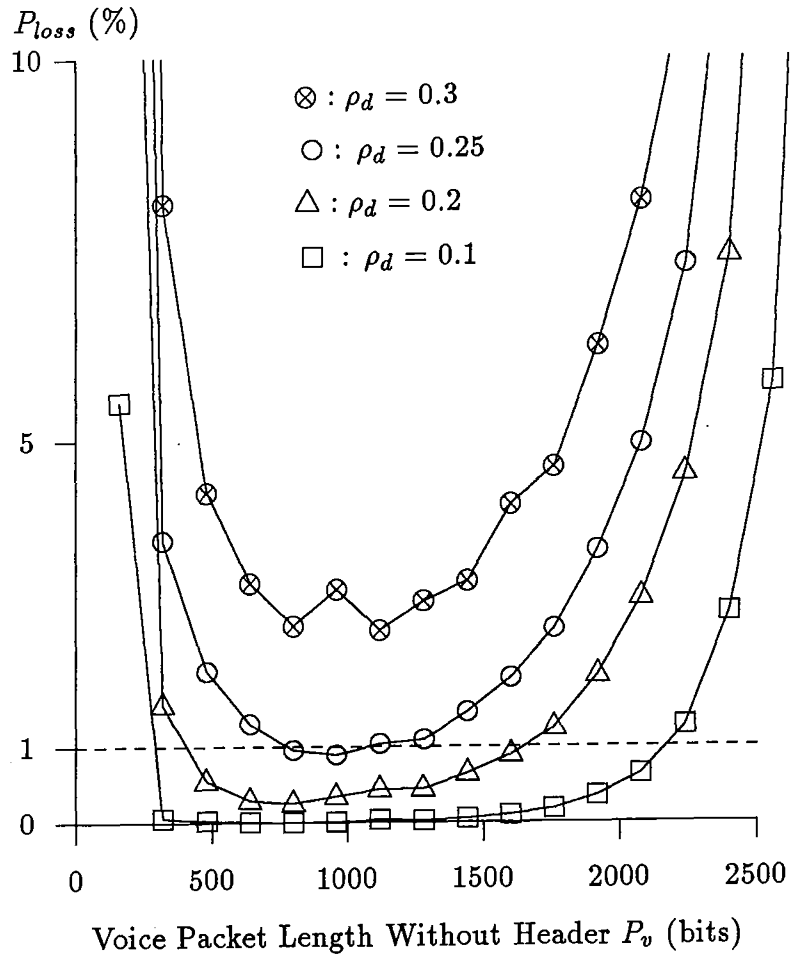

Fig.14 Voice Packet Loss Prob. P_{loss}
($N_v = N_d = 15$, $P_d = 8192$ bits, $\tau = 50$ msec)

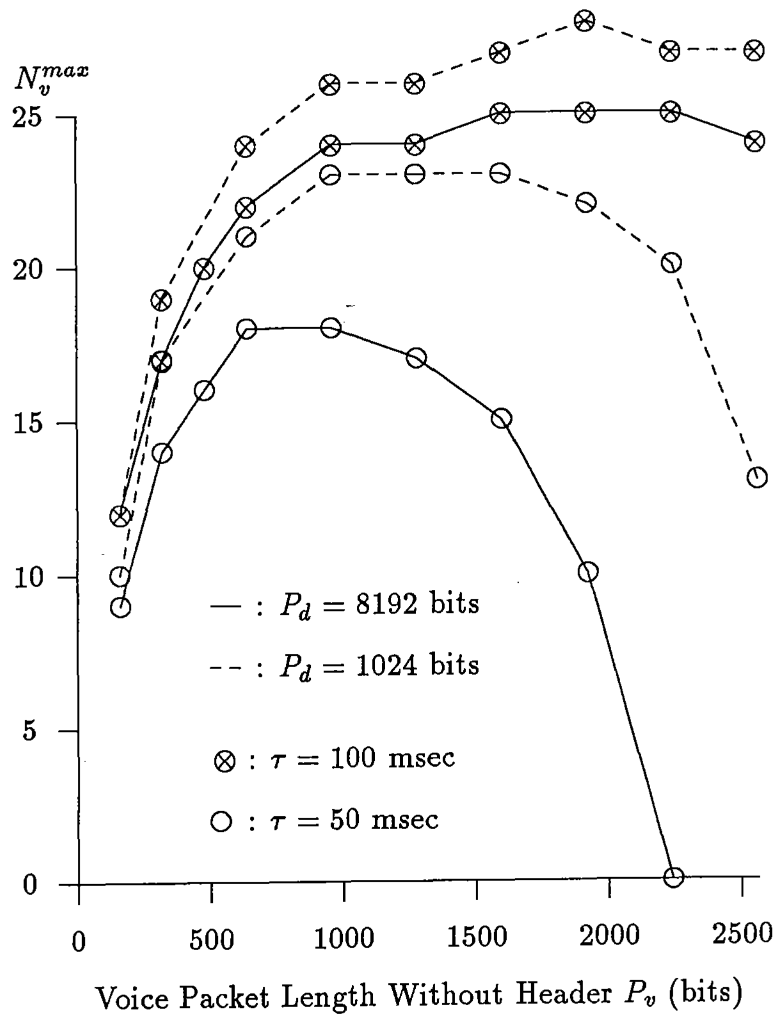

Fig.15 Maximum Number of Voice Calls Allowed on a Network
($N_d = 15$, $\rho_d = 0.2$)

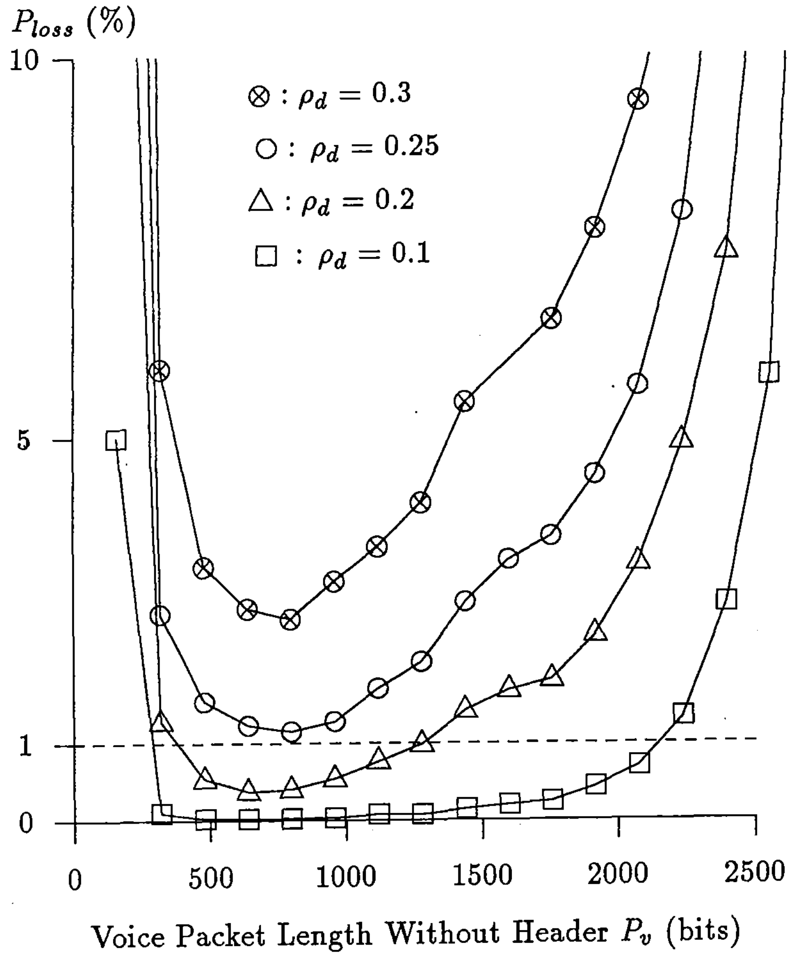

Fig.16 Voice Packet Loss Prob. P_{loss} in an Exhaustive Service Network
($N_v = N_d = 15$, $P_d = 8192$ bits, $\tau = 50$ msec)

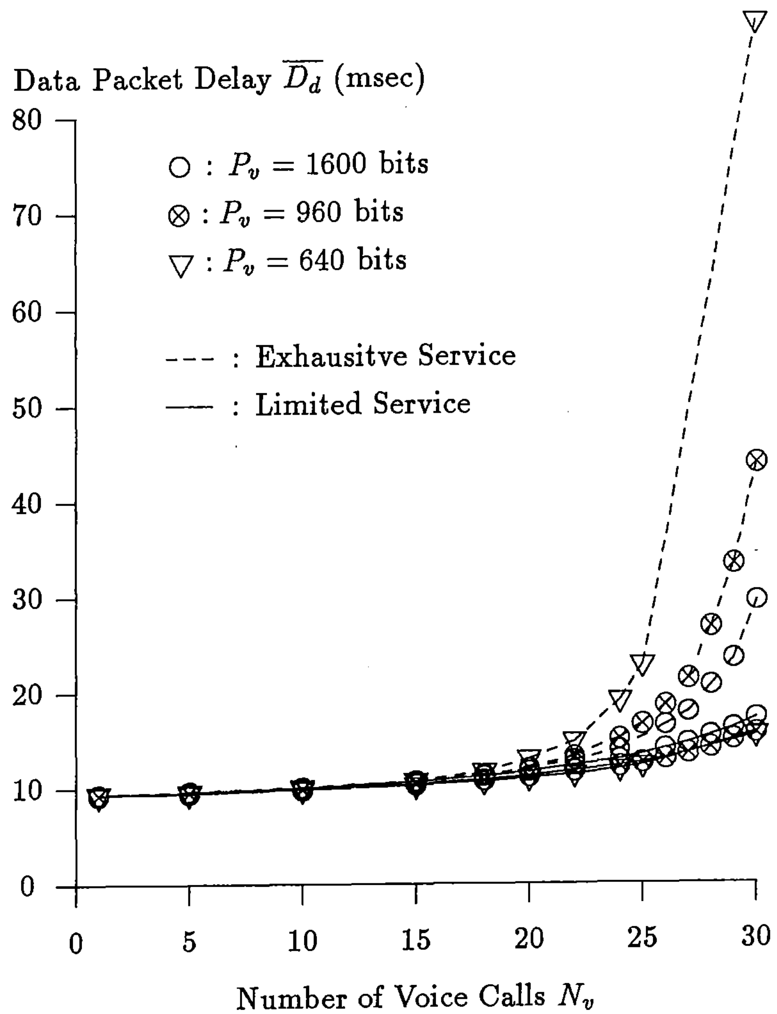

Fig.17 Data Packet Transmission Delay $\overline{D_d}$
($N_v = N_d = 15$, $P_d = 8192$ bits, $\tau = 50$ msec, $\rho_d = 0.2$)

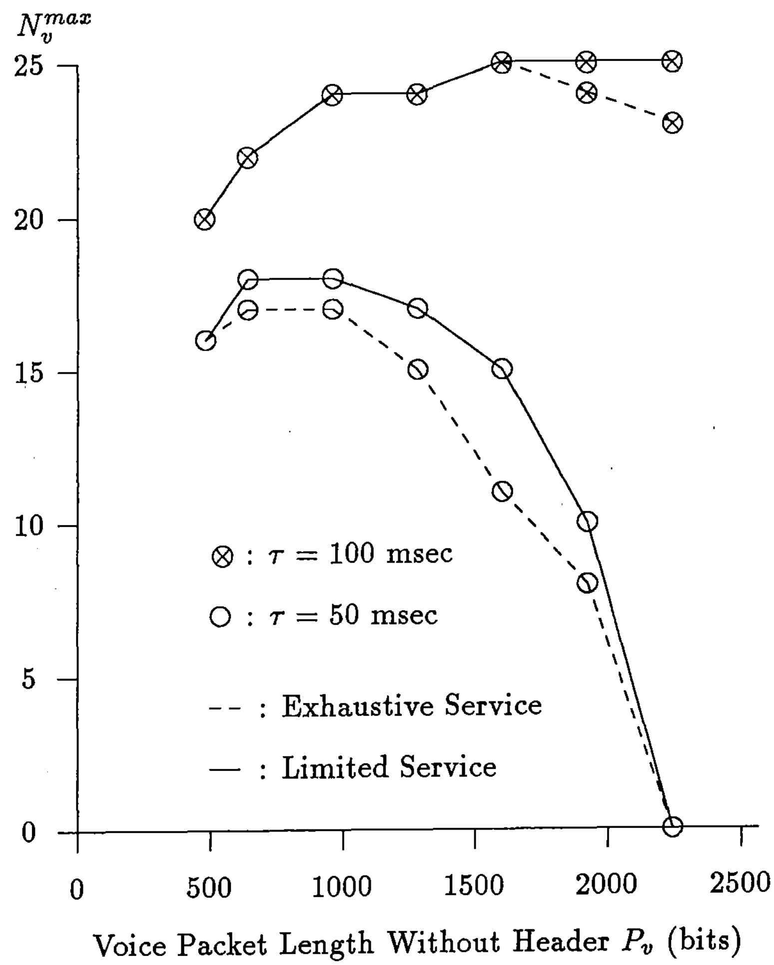

Fig.18 Maximum Number of Voice Calls Allowed on a Network
($N_d = 15$, $P_d = 8192$ bits, $\rho_d = 0.2$)